Defining gene and QTL networks

Jansen, Ritsert C.; Tesson, Bruno M.; Fu, Jingyuan; Yang, Yajie; McIntyre, Lauren M.; McInttyre, L.M.

Published in:
Current Opinion in Plant Biology

DOI:
10.1016/j.pbi.2009.01.003

Publication date:
2009

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Plant Biology

Defining gene and QTL networks

Ritsert C Jansen[1,2], Bruno M Tesson[1], Jingyuan Fu[1,2], Yajie Yang[3] and Lauren M McIntyre[3,4]

Current technologies for high-throughput molecular profiling of large numbers of genetically different individuals offer great potential for elucidating the genotype-to-phenotype relationship. Variation in molecular and phenotypic traits can be correlated to DNA sequence variation using the methods of quantitative trait locus (QTL) mapping. In addition, the correlation structure in the molecular and phenotypic traits can be informative for inferring the underlying molecular networks. For this, new methods are emerging to distinguish among causality, reactivity, or independence of traits based upon logic involving underlying QTL. These methods are becoming increasingly popular in plant genetic studies as well as in studies on many other organisms.

Addresses
[1] Groningen Bioinformatics Centre, University of Groningen, The Netherlands
[2] Department of Genetics, University Medical Centre Groningen, University of Groningen, The Netherlands
[3] Genetics Institute, University of Florida, USA
[4] Department of Molecular Genetics and Microbiology, University of Florida, USA

Corresponding author: McIntyre, Lauren M (mcintyre@ufl.edu)

Current Opinion in Plant Biology 2009, **12**:241–246

This review comes from a themed issue on
Genome studies and molecular genetics
Edited by Masahiro Yano and Roberto Tuberosa

Available online 3rd February 2009

1369-5266/$ – see front matter

DOI 10.1016/j.pbi.2009.01.003

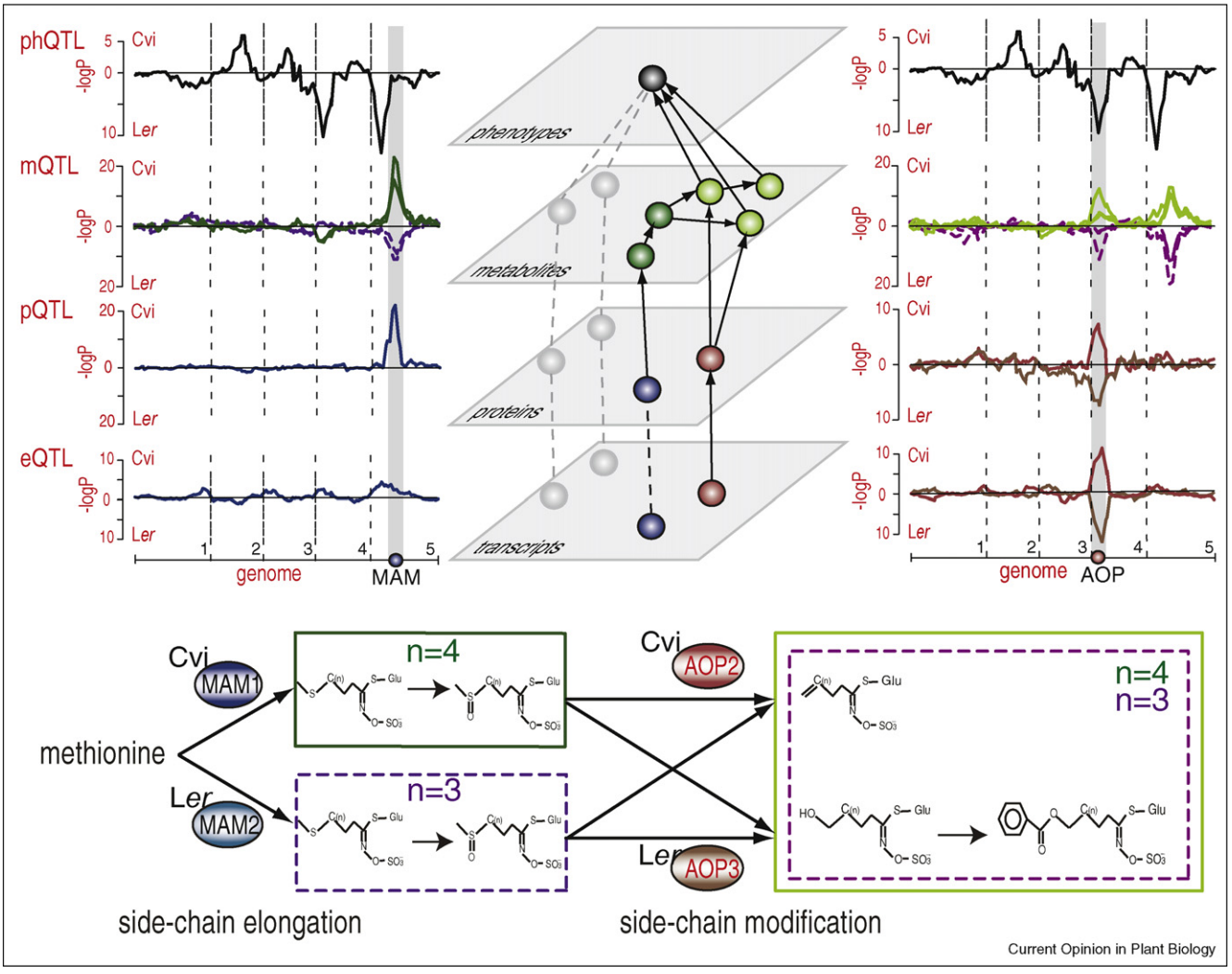
Introduction

Since the rediscovery of Gregor Mendel's pioneering work on pea crosses, segregating populations have been used to explore the underlying genetic architecture. The motivation for such work ranges from basic understanding of how life is organized and has adapted to changing environments, to utilization of such knowledge for genetic improvement of crops. Quantitative traits were deconstructed into additive, dominant, and epistatic effects, without consideration for the underlying molecular components. Technological advances in the 1980s made comprehensive genotyping affordable and mapping the rough location of the underlying genetic contribution for a quantitative trait locus (QTL) became feasible [1,2]. To date, more than 1200 studies in plants have been published on mapping phenotypic QTL (phQTL). A new wave of technological advances makes it possible to profile segregating populations for thousands of gene expression phenotypes and map expression QTL (eQTL) [3]. New technology can be used for the parallel measurement of the abundance of 1000s of proteins and metabolites to map protein QTL (pQTL) and metabolite QTL (mQTL). Deep sequencing, chromatin, and methyl-DNA immunoprecipitation are just a few of the newest technologies that add to the impressive arsenal of tools available for the study of the genetic variation underlying quantitative phenotypes [4,5]. Mapping phenotypes for thousands of traits, 'genetical genomics' [6••,7], is the first step in attempting to reconstruct gene networks. Methods for network reconstruction can be used within a particular level (intra-level analysis, i.e. transcript data only), to explain the relationship among traits [8•] at that level. Alternatively, the focus can be on understanding relationships across levels (inter-level analysis, integrating transcript, protein, metabolite, and morphological phenotypic data). Prior knowledge from other experiments can also be incorporated to further develop the picture of the network. Figure 1 illustrates the challenges that can be encountered with real data.

Causal, reactive, or independent?

The examination of pairwise correlation between traits, or principal components summaries of these traits, can lead to the hypothesis of a functional relationship if that correlation is high [8•,9••,10•,11–13]. Incorporating QTL information, allows the inference of a functional relationship if two traits share multiple QTL, something that it is unlikely to happen at random. Going beyond the detected QTL, the correlation between residuals among traits, after accounting for QTL effects, or correlations between traits conditional on other traits, is further evidence for a network connection. To infer directional effects, it is necessary to analyze the correlations among pairs of traits in detail. If trait T_1 maps to a subset of the QTL of trait T_2, then the common QTL can be taken as evidence for their network connection, while the distinct QTL can be used to infer the direction [6••,8•]. If traits T_1 and T_2 have common QTL, without QTL that are distinct, then the inference is complicated and further analysis is needed to discriminate pleiotropy from any of the possible orderings among traits (Box 1 highlights approaches from [8•,9••,10•]). Although these problems are 'modern' the groundwork for such analyses are evident in the earliest days of quantitative analysis [14].

Figure 1



System-wide QTL analysis for aliphatic glucosinolates. Data in this example are taken from [41] and demonstrate important network features that reconstruction methods should take into account. The colors in the QTL likelihood graphs (upper panel) and in the pathway (lower panel) correspond. The sign of the QTL effect is shown by plotting the QTL likelihood above the *x*-axis if the Cvi allele has higher average trait value or below the *x*-axis if the L*er* allele has higher average trait value. The vertical dashed lines indicate the chromosome borders and the physical gene positions are shown as stars and gray vertical bars. Glucosinolates are important secondary metabolites in plants and are well-known for their toxic effect on insects. The aliphatic glucosinolate biosynthesis pathway is summarized in the lower panel. The MAM genes are involved in side-chain elongation process: MAM1 mainly synthesizes C3-glucosinolates (in purple box) and MAM2 mainly synthesizes C4-glucosinolates (in green box). The parents Cvi and L*er* carry different MAM genes. Cvi contains two MAM1 genes and L*er* contains a functional MAM2 in addition to a truncated, nonfunctional MAM1 gene. The AOP genes are involved in the side-chain modification process: AOP2 and AOP3 generate different types of glucosinolates as described. Both AOP2 and AOP3 are present in L*er* and Cvi. But AOP2 is only expressed in Cvi and AOP3 is only expressed in L*er*. The top panel shows the QTL profiles at different levels (transcripts, proteins, metabolites, and disease trait) from a Cvi × L*er* recombinant inbred line population. To clearly demonstrate the QTL effects at different levels and along the pathway, the components are divided into two parts: the left part relates to the MAM1 gene and the metabolites produced by MAM synthesis (MAM2 is not measured); the right part refers to the AOP genes and the metabolites produced by the AOP synthesis. *cis*-eQTL is detected for AOP2 and AOP3; *cis*-pQTL for MAM1, AOP2, and AOP3; and mQTL for the various aliphatic glucosinolates. These QTL have the same or opposite sign of QTL effect. To demonstrate whether QTL at molecular levels can propagate to phenotypic level, molecular QTL (eQTL, pQTL, and mQTL) are also compared to the QTL of insect susceptibility (phQTL). The disease trait maps to ERECTA, a gene well-known for its widespread pleiotropic effect, to AOP, and to third gene, but it does not map to MAM. This can be explained by the fact that the total glucosinolate content maps to AOP only.

Intra-level analysis

In reference organisms, such as *Arabidopsis*, and in a growing list of plants, the location of the genes producing the transcript or protein studied is known. This added information provides a layer of interpretation for eQTL and pQTL. In *Arabidopsis*, eQTL and pQTL networks have been defined [15–21]; in barley, eucalyptus, and maize eQTL networks have been defined [22–27]. When

Box 1 Advanced causal reasoning

For traits T_1 and T_2, $T_1 > T_2$ denotes that T_1 affects T_2; $T_1 < T_2$ denotes that T_2 affects T_1, and T_1–T_2 denotes a correlation of unknown direction. If traits T_1 and T_2 have one common QTL, without QTL that are distinct, then the inference of causality is complicated and further analysis needed to discriminate pleiotropy from any of the possible orderings among traits. In this case there are at least three possible models: QTL > T_1 > T_2; QTL > T_2 > T_1; QTL > T_1 and QTL > T_2. If we write the simple regression models $T_2 = \alpha_2 + \beta_2 QTL + \varepsilon_2$ and $T_1 = \alpha_1 + \beta_1 QTL + \varepsilon_1$ and if ε_1 and ε_2 are uncorrelated, the QTL may be considered to have pleiotropic effects on the two traits, that is with no direct link between T_1 and T_2. Alternatively, if there is no evidence for pleiotropy, then the following models can be considered $T_2 = \alpha_3 + \beta_3 T_1 + \varepsilon_3$ and $T_1 = \alpha_4 + \beta_4 T_2 + \varepsilon_4$. The residuals from these models can be used to infer the correct model. If QTL > T_1 > T_2 is the true relation, then ε_3 will not map to the QTL. In contrast, ε_4 should have a residual signature of the QTL. Similarly, if QTL > T_2 > T_1 is the true relationship, then ε_4 will not map to the QTL, and ε_3 should have a residual signature of the QTL. In other cases, that is ε_3 and ε_4 together map or do not map to the QTL, the directionality is not clearly indicated [9••]. In addition, there are other competitive models such as QTL > T_1 > T_2 *and* QTL > T_2; or a loop QTL > T_1 > T_2 *and* T_2 > T_1 that prevent clear (conclusive) inferences about the true network directions [8•,10•]. As an important cautionary note, the above conditional models are based on various assumptions, and violation of these assumptions may lead to an increase in error rates for inferences about network structure.

the eQTL or pQTL colocalize with the gene, this effect may be due to *cis* regulatory effects (Figure 1). The caveat is that the detection of *cis* effects may be an artifact of differential probe hybridization because of sequence polymorphism [28,29]. If gene expression at a particular locus is regulated by that locus (*cis* effect), and the abundance of the transcript in turn regulates additional loci (*trans* effect) then these expression traits should all map to the same locus. If the number of *trans* loci regulated by a single locus is large, as would be expected from a master regulator, or switch, a *trans* band will be observed at this location. All genes in the QTL are candidate regulators; their partial correlations with the regulated genes can be used to prioritize them [30,31]. Importantly, genes without *cis*-eQTL can be regulators, manifesting only at the protein level [32]. If the number of transcript or protein traits mapping to a single location exceeds the number expected by chance, then a hotspot has been identified [33,34]. The hotspot can be inferred to represent a possible master regulator or switch. However, as a cautionary note hotspots can be an artifact of improper permutation [34].

At the metabolite level, mQTL for traits connected in a network may show complex patterns of correlation. For example, the mQTL for the precursor and product of an enzymatic step with differential activity should have opposite signs — indicating that the sign of the mQTL effect also conveys valuable information [35–37]. The effect of an mQTL may be visible on the precursor, the corresponding product and downstream products (Figure 1). As the number of steps grows, the complexity of the network increases, and network reconstruction based purely on correlation coefficients is challenging. Epistatic interactions among enzymes may further complicate the effort to map and deconstruct their unique patterns, as in the cases where some allelic combinations can be found in offspring which will then produce metabolites not found in either parent. Although such epistasis may be a rare phenomenon of complex traits [38], it is potentially abundant in secondary metabolism [36,37].

Inter-level analysis

Inter-level inferences have been made between eQTL and mQTL in *Arabidopsis* [39] and between mQTL and phenotypic traits in tomato [40], and between eQTL, pQTL, mQTL, and phenotypic traits in *Arabidopsis* [41]. A system-wide analysis can reveal the impact of DNA sequence variation across multiple levels, that is eQTL at the gene expression level, pQTL for protein abundance or activity traits, mQTL for metabolite abundances and/or phQTL for morphological traits (Figure 1). Some DNA sequence variation will induce strong effects to be detected as hotspots or master regulators of many molecular and phenotypic traits, while others induce effects that are more subtle or are buffered in the network to ensure robustness of the system [41]. Correlations among traits from different levels can be used to generate hypotheses about network connections in inter-level analyses. Principal components may be used to summarize a network on one level and then regressed on traits on another level [42]. The complexity of the system is such that two adjacent levels (i.e. transcript and protein) may not be linearly related. For example, DNA sequence variation may not affect expression level (no eQTL) while it does affect protein abundance or activity (pQTL). The 'higher' level traits (phQTL) may also be a function of multiple underlying (perhaps interacting) subnetworks (see the disease trait in Figure 1). Added complexity may be observed when DNA sequence variation directly affects higher level traits that — through feedback loops — affect other traits at the same or lower levels [39]. These examples indicate that caution is warranted given the intrinsic complexity in real networks.

Correlation analyses will only reveal the linear relationships among levels. Interpreting the correlation structure 'beyond' the common and shared QTL, using methods such as those described in Box 1, may generate hypotheses about system-wide networks. However, extreme caution is advisable in these interpretations in intra-level analyses owing to the potential impact of correlated measurement error (leading to false positive connections), and in inter-level analyses owing to the seeming lack of correlation of between levels (leading to false negative connections) [43].

Using *a priori* knowledge

Structural and functional data (gene sequence, gene localization, transcription factor binding sites (TFBS), Gene Ontology (GO), metabolic pathway, and protein–protein interaction (PPI)) as well as independent experimental data gleaned from secondary sources (i.e. Gene Expression Omnibus (GEO)) can be used post hoc to verify the defined gene and QTL networks. For example, if a disease maps to multiple QTL, then the candidate genes in each of the QTL can be analyzed and prioritized using known functional interactions [44]. As another example, particular eQTL *trans* bands may be identified as significantly enriched for a functional GO category [45] or as more likely to represent binding sites for transcription factors [46]. Prior knowledge can also be integrated in analysis. For example, a set of pathway-related genes may not show significant eQTL in gene-by-gene tests, while the set of genes can show such significance in a group-wise test [18].

Future directions

Genetic variation at multiple loci in combination with environmental factors can induce molecular or phenotypic variation. Variation may manifest itself as linear patterns among traits at different levels that can be deconstructed. Correlations can be attributed to detectable QTL and a logical framework based on common and distinct QTL can be used to infer network causality, reactivity, or independence. Unexplained variation can be used to infer direction between traits that share a common QTL and have no distinct QTL. Unexplained variation originates from other minor or modifier QTL, epigenetic factors, and biological, environmental and unfortunately, technical factors. Correlation structures present in the molecular data may reflect technical artifacts, in which case the models used to infer causality are potentially invalid and the inference is potentially erroneous. Additional studies are needed to understand and quantify the level of sensitivity of these network reconstruction methods to technical errors. Further research is also needed to develop and evaluate experimental designs other than the current biparental line crosses: for example, multiple line crosses [47–49], advanced intercrosses [47,48], or populations of natural ecotypes [49–54]. Prior knowledge and complementary experiments such as deletion mapping followed by independent gene expression studies between parental lines may validate or disprove implicated network connections [55••].

The trend of genetic studies to go deeper (more levels) and broader (larger scale and more factors, including environmental ones) brings challenges to develop methodology that can reconstruct networks more efficiently and more accurately. Despite the obvious limitations of gene and QTL network reconstruction methods, these and other future developments in biotechnology and genetics hold for sure great promise for the field of quantitative genetics.

Acknowledgment

Netherlands Organization for Scientific Research (NWO) and NSF (0821954).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Doerge RW: **Mapping and analysis of quantitative trait loci in experimental populations**. *Nat Rev Genet* 2002, **3**:43-52.

2. Jansen RC: **Studying complex biological systems using multifactorial perturbation**. *Nat Rev Genet* 2003, **4**:145-151.

3. Gilad Y, Rifkin SA, Pritchard JK: **Revealing the architecture of gene regulation: the promise of eQTL studies**. *Trends Genet* 2008, **24**:408-415.

4. Johannes F, Colot V, Jansen RC: **Epigenome dynamics: a quantitative genetics perspective**. *Nat Rev Genet* 2008, **9**:883-890.

5. Martienssen RA, Doerge RW, Colot V: **Epigenomic mapping in *Arabidopsis* using tiling microarrays**. *Chromosome Res* 2005, **13**:299-308.

6. •• Jansen RC, Nap JP: **Genetical genomics: the added value from segregation**. *Trends Genet* 2001, **17**:388-391.
This paper introduces the name, concept, and potential of 'genetical genomics' for network reconstruction.

7. Li Y, Breitling R, Jansen RC: **Generalizing genetical genomics: getting added value from environmental perturbation**. *Trends Genet* 2008, **24**:518-524.

8. • Chaibub Neto E, Ferrara CT, Attie AD, Yandell BS: **Inferring causal phenotype networks from segregating populations**. *Genetics* 2008, **179**:1089-1100.
This paper provides the statistical framework for causal inference on the basis of common and distinct QTL. The relationship between the graphical network and the statistical model is elucidated.

9. •• Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C *et al.*: **An integrative genomics approach to infer causal associations between gene expression and disease**. *Nat Genet* 2005, **37**:710-717.
This paper describes causal reasoning in QTL studies.

10. • Li R, Tsaih SW, Shockley K, Stylianou IM, Wergedal J, Paigen B, Churchill GA: **Structural model analysis of multiple quantitative traits**. *PLoS Genet* 2006, **2**:e114.
This paper clearly enumerates all the possible causal relationships between a QTL and two phenotypes.

11. Adewale AJ, Dinu I, Potter JD, Liu Q, Yasui Y: **Pathway analysis of microarray data via regression**. *J Comput Biol* 2008, **15**:269-277.

12. Liu B, de la Fuente A, Hoeschele I: **Gene network inference via structural equation modeling in genetical genomics experiments**. *Genetics* 2008, **178**:1763-1776.

13. Zhu J, Wiener MC, Zhang C, Fridman A, Minch E, Lum PY, Sachs JR, Schadt EE: **Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations**. *PLoS Comput Biol* 2007, **3**:e69.

14. Wright S: **Correlation and causation**. *J Agric Res* 1921, **201**:557-585.

15. DeCook R, Lall S, Nettleton D, Howell SH: **Genetic regulation of gene expression during shoot development in *Arabidopsis***. *Genetics* 2006, **172**:1155-1164.

16. Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJ, Vreugdenhil D, Koornneef M, Jansen RC: **Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci**. *Proc Natl Acad Sci U S A* 2007, **104**:1708-1713.

17. Wang D, Nettleton D: **Combining classical trait and microarray data to dissect transcriptional regulation: a case study**. *Theor Appl Genet* 2008, **116**:683-690.

18. Kliebenstein DJ, West MA, van Leeuwen H, Loudet O, Doerge RW, St Clair DA: **Identification of QTLs controlling gene expression networks defined a priori**. *BMC Bioinformatics* 2006, **7**:308.

19. Juenger TE, Wayne T, Boles S, Symonds VV, McKay J, Coughlan SJ: **Natural genetic variation in whole-genome expression in *Arabidopsis thaliana*: the impact of physiological QTL introgression**. *Mol Ecol* 2006, **15**: 1351-1365.

20. West MA, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW, St Clair DA: **Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis***. *Genetics* 2007, **175**:1441-1450.

21. Vuylsteke M, Daele H, Vercauteren A, Zabeau M, Kuiper M: **Genetic dissection of transcriptional regulation by cDNA-AFLP**. *Plant J* 2006, **45**:439-446.

22. Druka A, Potokina E, Luo Z, Bonar N, Druka I, Zhang L, Marshall DF, Steffenson BJ, Close TJ, Wise RP *et al.*: **Exploiting regulatory variation to identify genes underlying quantitative resistance to the wheat stem rust pathogen *Puccinia graminis* f. sp. tritici in barley**. *Theor Appl Genet* 2008, **117**:261-272.

23. Kirst M, Basten CJ, Myburg AA, Zeng ZB, Sederoff RR: **Genetic architecture of transcript-level variation in differentiating xylem of a eucalyptus hybrid**. *Genetics* 2005, **169**:2295-2303.

24. Potokina E, Druka A, Luo Z, Moscou M, Wise R, Waugh R, Kearsey M: **Tissue-dependent limited pleiotropy affects gene expression in barley**. *Plant J* 2008, **56**:287-296.

25. Potokina E, Druka A, Luo Z, Wise R, Waugh R, Kearsey M: **Gene expression quantitative trait locus analysis of 16,000 barley genes reveals a complex pattern of genome-wide transcriptional regulation**. *Plant J* 2008, **53**:90-101.

26. Shi C, Uzarowska A, Ouzunova M, Landbeck M, Wenzel G, Lubberstedt T: **Identification of candidate genes associated with cell wall digestibility and eQTL (expression quantitative trait loci) analysis in a Flint × Flint maize recombinant inbred line population**. *BMC Genomics* 2007, **8**:22.

27. Springer NM, Stupar RM: **Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize**. *Plant Cell* 2007, **19**:2391-2402.

28. Gilad Y, Borevitz J: **Using DNA microarrays to study natural variation**. *Curr Opin Genet Dev* 2006, **16**:553-558.

29. Alberts R, Terpstra P, Li Y, Breitling R, Nap JP, Jansen RC: **Sequence polymorphisms cause many false *cis* eQTLs**. *PLoS ONE* 2007, **2**:e622.

30. Bing N, Hoeschele I: **Genetical genomics analysis of a yeast segregant population for transcription network inference**. *Genetics* 2005, **170**:533-542.

31. Kulp DC, Jagalur M: **Causal inference of regulator–target pairs by gene mapping of expression phenotypes**. *BMC Genomics* 2006, **7**:125.

32. Stylianou IM, Affourtit JP, Shockley KR, Wilpan RY, Abdi FA, Bhardwaj S, Rollins J, Churchill GA, Paigen B: **Applying gene expression, proteomics and single-nucleotide polymorphism analysis for complex trait gene identification**. *Genetics* 2008, **178**:1795-1805.

33. de Koning DJ, Haley CS: **Genetical genomics in humans and model organisms**. *Trends Genet* 2005, **21**:377-381.

34. Breitling R, Li Y, Tesson BM, Fu J, Wu C, Wiltshire T, Gerrits A, Bystrykh LV, de Haan G, Su AI *et al.*: **Genetical genomics: spotlight on QTL hotspots**. *PLoS Genet* 2008, **4**:e1000232.

35. Fu J, Swertz MA, Keurentjes JJ, Jansen RC: **MetaNetwork: a computational protocol for the genetic study of metabolic networks**. *Nat Protoc* 2007, **2**:685-694.

36. Keurentjes JJ, Fu J, de Vos CH, Lommen A, Hall RD, Bino RJ, van der Plas LH, Jansen RC, Vreugdenhil D, Koornneef M: **The genetics of plant metabolism**. *Nat Genet* 2006, **38**:842-849.

37. Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ: **Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome**. *Plant Cell* 2008, **20**:1199-1216.

38. Hill WG, Goddard ME, Visscher PM: **Data and theory point to mainly additive genetic variance for complex traits**. *PLoS Genet* 2008, **4**:e1000008.

39. Wentzell AM, Rowe HC, Hansen BG, Ticconi C, Halkier BA, Kliebenstein DJ: **Linking metabolic QTLs with network and *cis*-eQTLs controlling biosynthetic pathways**. *PLoS Genet* 2007, **3**:1687-1701.

40. Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR: **Mode of inheritance of primary metabolic traits in tomato**. *Plant Cell* 2008, **20**:509-523.

41. Fu J, Keurentjes JJB, Bouwmeester H, America T, Verstappen FWA, Ward JL, Beale MH, de Vos RCH, Dijkstra M, Scheltema RA, et al.: **System-wide molecular evidence for phenotypic buffering in Arabidopsis**. *Nat Genet* 2009, *in press*, doi:10.1038/ng.308.

42. Coffman CJ, Wayne ML, Nuzhdin SV, Higgins LA, McIntyre LM: **Identification of co-regulated transcripts affecting male body size in *Drosophila***. *Genome Biol* 2005, **6**:R53.

43. Jansen RC, Nap JP, Mlynarova L: **Errors in genomics and proteomics**. *Nat Biotechnol* 2002, **20**:19.

44. Franke L, van Bakel H, Fokkens L, de Jong ED, Egmont-Petersen M, Wijmenga C: **Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes**. *Am J Hum Genet* 2006, **78**: 1011-1025.

45. Wu C, Delano DL, Mitro N, Su SV, Janes J, McClurg P, Batalov S, Welch GL, Zhang J, Orth AP *et al.*: **Gene set enrichment in eQTL data identifies novel annotations and pathway regulators**. *PLoS Genet* 2008, **4**:e1000070.

46. Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, Bumgarner RE, Schadt EE: **Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks**. *Nat Genet* 2008, **40**:854-861.

47. Chesler EJ, Miller DR, Branstetter LR, Galloway LD, Jackson BL, Philip VM, Voy BH, Culiat CT, Threadgill DW, Williams RW *et al.*: **The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics**. *Mamm Genome* 2008, **19**:382-389.

48. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W *et al.*: **The Collaborative Cross, a community resource for the genetic analysis of complex traits**. *Nat Genet* 2004, **36**:1133-1137.

49. Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C *et al.*: **Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes**. *PLoS Genet* 2005, **1**:e60.

50. Gonzalez-Martinez SC, Huber D, Ersoz E, Davis JM, Neale DB: **Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination**. *Heredity* 2008, **101**:19-26.

51. Kim S, Zhao K, Jiang R, Molitor J, Borevitz JO, Nordborg M, Marjoram P: **Association mapping with single-feature polymorphisms**. *Genetics* 2006, **173**:1125-1133.

52. Rosenberg NA, Nordborg M: **A general population-genetic model for the production by population structure of spurious genotype–phenotype associations in discrete, admixed or spatially distributed populations**. *Genetics* 2006, **173**:1665-1678.

53. Zhao J, Paulo MJ, Jamar D, Lou P, van Eeuwijk F, Bonnema G, Vreugdenhil D, Koornneef M: **Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa***. *Genome* 2007, **50**:963-973.

54. Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P *et al.*: **An *Arabidopsis* example of association mapping in structured samples**. *PLoS Genet* 2007, **3**:e4.

55. Wayne ML, McIntyre LM: **Combining mapping and arraying: an •• approach to candidate gene identification**. *Proc Natl Acad Sci U S A* 2002, **99**:14903-14906.
This is the first paper to show that eQTL data can be used to prioritize genes for phQTL.